DOCUMENT RESUME

ED 389 744 TM 024 377

AUTHOR	Feuer, Michael J., Ed.; Kober, Nancy, Ed.
TITLE	Anticipating Goals 2000. Standards, Assessment, and Public Policy. Summary of a Workshop (Washington, D.C., March 9, 1994). Board Bulletin.
INSTITUTION	National Academy of Sciences - National Research Council, Washington, DC. Commission on Behavioral and Social Sciences and Education.
PUB DATE	95
NOTE	36p.
AVAILABLE FROM	Board of Testing and Assessment, National Research Council, 2101 Constitution Avenue N.W., Washington, DC 20418.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Standards; *Accountability; *Educational Assessment; Educational Change; Educational Policy; Elementary Secondary Education; Policy Formation; *Public Policy; School Districts; State Programs; *Test Construction; Workshops
IDENTIFIERS	Goals 2000; Reform Efforts

ABSTRACT

The Board on Testing and Assessment of the National Research Council convened a workshop in 1994 to help policymakers and others better understand the complex issues emerging from the standards-based educational reform movement. This bulletin synthesizes the workshop discussions. It is organized around four major themes that emerged from the presentations: (1) the implications of using standards as accountability tools; (2) the challenges of designing assessments related to standards; (3) the implications of building the new form of education federalism implied by standards-based reform; and (4) the challenges of strengthening the state and local capacities to implement standards and linked assessments. Each section contains a brief review of the main issue, a synthesis of the views raised during the workshop discussion, and a list of questions for further analysis. Steps that will have to be taken to make the vision of standards-based reform a reality are outlined. An appendix describes the workshop agenda and lists participants. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

BOARD BULLETIN

ED 389 744

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
W.M. BURNS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Anticipating Goals 2000

Standards, Assessment, and Public Policy

Summary of a Workshop

TM024377

BEST COPY AVAILABLE

2

NATIONAL RESEARCH COUNCIL

BOARD BULLETIN

Anticipating Goals 2000

Standards, Assessment, and Public Policy

Summary of a Workshop

Michael J. Feuer and Nancy Kober, editors

Board on Testing and Assessment
Commission on Behavioral and Social Sciences and Education
National Research Council

National Academy Press
Washington, D.C. 1995

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

The work of the Board on Testing and Assessment is supported by the U.S. Departments of Defense, Education, and Labor, through a grant administered by the Employment and Training Administration of the U.S. Department of Labor.

Additional copies of this report are available from:
Board on Testing and Assessment
National Research Council
2101 Constitution Avenue N.W.
Washington, D.C. 20418

Printed in the United States of America
Copyright 1995 by the National Academy of Sciences. All rights reserved.

BOARD ON TESTING AND ASSESSMENT

Richard C. Atkinson (*Chair*), University of California, San Diego

Constance B. Newman (*Vice Chair*), Smithsonian Institution, Washington, D.C.

Richard J. Shavelson (*Vice Chair*), School of Education, Stanford University

Laurie J. Bassi, Graduate Public Policy Program, Georgetown University

David C. Berliner, College of Education, Arizona State University, Tempe

Richard F. Elmore, Graduate School of Education, Harvard University

Patricia M. Flynn, Graduate School of Business, Bentley College

Edmund W. Gordon, Department of Psychology, City University of New York

Sylvia T. Johnson, School of Education, Howard University

Brigitte Jordan, Xerox Palo Alto Research Center and Institute for Research on Learning, Palo Alto, Calif.

Carl F. Kaestle, Department of Education, University of Chicago

Luis M. Laosa, Educational Testing Service, Princeton, N.J.

Renee S. Lerche, Ford Motor Company, Dearborn, Mich.

Alan M. Lesgold, Learning Research and Development Center, University of Pittsburgh

Robert L. Linn, School of Education, University of Colorado, Boulder

Miles A. Myers, National Council of Teachers of English, Urbana, Ill.

James L. Outtz, Outtz and Associates, Washington, D.C.

Neal W. Schmitt, Department of Psychology, Michigan State University, East Lansing

Alan H. Schoenfeld, School of Education, University of California, Berkeley

David S. Tatel,* Hogan and Hartson, Washington, D.C.

Ewart A.C. Thomas, Department of Psychology, Stanford University

Michael J. Feuer, *Director*

Holly Wells, *Administrative Assistant*

*Member, July 1993-October 1994

FOREWORD

Few activities on the public agenda have as much long-term significance for the health and prosperity of American democracy as a sustained commitment to improvement of education and the life-long development of our precious human resources.

The recent passage of the Goals 2000: Educate America Act and the reauthorization of the Elementary and Secondary Education Act are watershed events in American education history. The principle that all children can learn to high standards is now the law of the land, and a new partnership has been forged between the federal government and the states and local school districts to help that vision become a reality. Core elements in the new partnership are voluntary standards of content, performance, and opportunity to learn, as well as new approaches to assessment of student achievement that will provide a stimulus for—and benchmarks of—continued educational progress. The vision of systemic change embodied in the Goals 2000 legislation now requires careful attention to the details of implementation: How will standards be set? How will assessments be designed? What will be the effects on children and teachers? How can standards and assessments become effective tools of learning, teaching, and system accountability?

These and other questions are the subject of this bulletin, the first in a series anticipated by the Board on Testing and Assessment. The board is a relatively new entity of the National Research Council, which through its many committees and boards is deeply involved in applying scientific knowledge to education reform. As part of this commitment to improving education, the Board on Testing and Assessment will provide a scientific forum for increasing the understanding of the complex issues tied to standards, testing, and the assessment of human performance.

Bruce Alberts

Bruce Alberts, *Chair*
National Research Council

ACKNOWLEDGMENTS

The Board on Testing and Assessment is grateful to the many individuals whose efforts made this bulletin, and the workshop it summarizes, possible. The board's work is supported by generous grants from the U.S. Departments of Defense, Education, and Labor. The continued interest and encouragement offered by Steve Sellman, Jane Arabian, Emerson Elliot, Gary Phillips, Alan Ginsburg, Valena Plisko, Raymond Uhalde, Robert Litman, and Donna Dye are very much appreciated.

We also wish to thank the workshop presenters, whose remarks stimulated a rich and wide-ranging discussion: Gordon Ambach, Susan Fuhrman, Michael Kean, Dan Koretz, Shirley Malcom, and Phyllis McClure. During each segment of the workshop, the discussion was much enhanced by the insightful comments of board members who acted as discussants: David Berliner, Richard Elmore, Edmund Gordon, Sylvia Johnson, Alan Lesgold, Robert Linn, Alan Schoenfeld, David Tatel, and Ewart Thomas. Constance Newman, vice chair of the board, skillfully and gracefully guided the entire day's discussions.

The workshop and this bulletin were conceived and executed by Michael Feuer, staff director of the board. In helping to translate the day's proceedings into this summary, the work of Nancy Kober was exemplary. Several other National Research Council staff members read early drafts and provided helpful critiques: in particular, we thank Donna Gerardi, Linda Rosen, and Alexandra Wigdor. In addition, we wish to thank Steve Baldwin and Ray Fields for their careful reading and invaluable suggestions.

Special words of thanks go to Christine McShane for her fine-tuning, to Leigh Coriale for her creative design, and to Eugenia Grohman for patiently guiding us through the review and publication process. Finally, we thank Holly Wells for her excellent administrative support.

CONTENTS

USING STANDARDS AS ACCOUNTABILITY TOOLS ◦ 2

Inputs Versus Outcomes ◦ 3

Challenges of Developing Effective Standards ◦ 5

Legal Ramifications of Standards ◦ 7

DEVELOPING ASSESSMENTS ALIGNED WITH STANDARDS ◦ 9

Approaching Assessment Development ◦ 10

Lessons From Vermont ◦ 11

Technical Questions ◦ 12

Appropriate Use of Assessments ◦ 13

Effects on Teaching and Learning ◦ 14

A NEW ERA OF EDUCATION FEDERALISM ◦ 15

Title I: A Major Influence ◦ 16

Certifying Standards and Assessments ◦ 17

STRENGTHENING STATE AND LOCAL CAPACITY ◦ 18

New Governance Arrangements ◦ 19

NEXT STEPS ◦ 21

APPENDIX: WORKSHOP AGENDA AND PARTICIPANTS ◦ 23

Anticipating Goals 2000

Standards, Assessment, and Public Policy

Anticipating Goals 2000:

Standards, Assessment, and Public Policy

Encouraged by the *Goals 2000: Educate America Act* and other federal and state legislation, a movement is under way to reform education by establishing ambitious standards at the national and state levels to guide the content of learning in core subjects, the performance expectations for all students, and the opportunities to learn afforded all children. Important components of this strategy are assessments aimed at measuring the progress of students, schools, districts, and states toward the achievement of the content standards. This shift toward voluntary national goals, standards, and assessments is a watershed in American education history and will influence the course of public schooling for years to come.

Because standards-based reform could have enormous consequences for the ways millions of American schoolchildren are taught and assessed, as well as for the ways in which millions of young Americans are prepared and selected for productive employment, it is critical to explore the many educational, social, technological, and political dimensions of the reform strategy. The Board on Testing and Assessment (BOTA) believes that among its principal roles are to elucidate the underlying assumptions and expected effects of reforms that rely heavily on standards and assessment and to provide objective and scientifically rigorous information to policy makers charged with its implementation.

∘ ∘ ∘ ∘

"For the first time in the nation's history, we have codified in federal law a set of national educational goals, along with the concept of voluntary national standards for all students."

Constance B. Newman

∘ ∘ ∘ ∘

Toward this end, BOTA convened a day-long workshop on March 9, 1994, at the National Academy of Sciences in Washington, D.C. The goals of the workshop were to help policy makers and others better understand the complex issues emerging from the standards-based reform movement, to elevate the level of discourse about standards and assessments beyond conventional wisdom and common generalities, and to highlight areas in which further research and exploration are needed.

The workshop format included presentations on critical national and state issues con-

cerning standards and assessments, responses from specific board members, and a frank, free-ranging exchange among the entire board and invited participants. Observers included federal agency officials, congressional staff, representatives of professional associations and standards-setting bodies, state and local educators, education researchers, scientists, and others. This bulletin, the first of several publications intended to acquaint a wide audience with BOTA activities and deliberations, synthesizes the proceedings of the March 9 workshop. It is important to note that, as a workshop summary, this document is limited in its scope by the discussions that actually took place. At the same time, we have attempted to draw attention to certain issues that the board considers important to its current and future work. The bulletin is organized around four major themes that emerged from the presentations and that were discussed throughout the day:

- the implications of using standards as accountability tools;
- the challenges of designing assessments related to standards;
- the implications of building the new form of education federalism implied by standards-based reform; and
- the challenges of strengthening the state and local capacity to implement standards and linked assessments.

Following are four sections exploring each of these themes. Within each section is a brief review of the main issue, a synthesis of views raised during the workshop discussion, and a list of questions for further analysis.

USING STANDARDS AS ACCOUNTABILITY TOOLS

THE ISSUE IN BRIEF

Advocates of standards-based reform argue that methods traditionally used by states and the federal government to instill accountability—namely, regulating "inputs" into schools, such as minimum resource and process requirements—have not worked very well to ensure educational quality. It is time, they say, to loosen the federal and state input requirements that have locked up the system, in exchange for greater attention to outcome standards specifying the content knowledge and skills to be taught and learned and the levels of performance to be attained. In this way, accountability for tax dollars would be enforced by assessing, monitoring, publicizing outcomes, and possibly attaching sanctions and rewards to performance.

Some contend that it is unfair to hold students, teachers, or school authorities accountable to content and performance standards without also defining the conditions that must exist in schools to afford students the opportunities to meet performance expectations—hence the emphasis on "opportunity-to-learn" standards.

How accurate are conventional notions about the effects of input and outcome requirements? Are there other ways to analyze or predict the effects of input and outcome requirements? What challenges must be addressed to make standards effective tools for institutional accountability and student motivation? What are the legal ramifications of new accountability approaches? These questions dominated much of the workshop discussion.

VIEWS FROM THE WORKSHOP

Inputs Versus Outcomes

An important theme in the discussion was that, in designing a system of accountability based on standards, it helps to move away from oversimplified notions such as (1) that more inputs necessarily lead to better results or (2) that results can be improved without consideration of the possible need for increased inputs. Instead, the workshop discussion focused on relationships between inputs and outcomes. One misleading notion is that input and outcome requirements are polar opposites. Rather, some fundamental beliefs appear to be shared by those who favor an emphasis on outcomes and those who advocate regulation of inputs. For example, those who support fiscal incentives to schools with outstanding performance must implicitly assume that extra monetary inputs make a difference, else they would have little value as rewards. And those who explicitly urge continued attention to the inputs side must presume that inputs will ultimately produce tangible outcomes for students.

° ° ° °

"People who believe in the efficacy of performance incentives are no different in their assumptions from people who believe in the efficacy of inputs."

Ewart Thomas

° ° ° °

Another common metaphor views input requirements and outcome standards as substitutable tools to enhance performance: performance can be enhanced by either raising (or setting) higher outcome standards or by raising (or setting) input levels. Yet, it was argued, the trade-offs are not so clean. Input requirements play an important and

° ° ° °

"We need to review existing regulation, not with the notion that we're going to eliminate it, but with the notion that we're going to ration it, streamline it,... and focus on regulation that makes enforcement and compliance likely."

Susan Fuhrman

° ° ° °

"Educators are saying that we don't know what the production function is: there is a great deal of uncertainty about the relationship between inputs and outputs. We might devote some intellectual resources to looking at how variations in inputs lead to very different worlds."

Laurie Bassi

° ° ° °

necessary role, even in an outcome-oriented governance system. Some desirable results of schooling cannot be captured very well by outcome measures. Moreover, there will probably always be school districts that will not provide the necessary inputs and equity guarantees unless directed to do so. In fact, it was noted, there is a perverse logic in rewarding with deregulation those schools that have been successful under the current system. These observations suggest that, rather than asking how to balance input and outcome requirements, it may be more useful to ask which *combinations* of input and outcome policies are likely to ensure higher performance and equitable access to learning.

If opportunity-to-learn standards are going to be more than a replay of past experience or another layer of regulation, it may be advisable for states to look beyond the strategies used in the past—primarily centrally imposed mandates and incentives—toward more participatory strategies. Several options for doing so are described below in the discussion of strengthening state and local capacity.

The workshop discussion of inputs and outcomes turned also to the debate over the the usefulness of trying to determine a production function for education: to identify, or even quantify, which kinds of inputs produce particular kinds of student outcomes and then build those characteristics into standards and linked assessments. Is it possible, for instance, to identify how much training in specific content a teacher needs in order to teach students to a particular performance level?

Some observers assert that classrooms are too idiosyncratic and education too much of a human enterprise to be quantified in this way. Yet for standards-based reform to work, it was noted, we must reach some conclusions about what kinds of instructional strategies, professional development, and organizational policies lead to higher outcomes—whether or not we call this a production function.[1]

1 One participant offered this suggestion after the workshop: "The metaphor of a recipe may be better than the black box of the production function. Not only do we need ingredients (books, curricula, teachers) but we also need to know how to cook the dish, i.e., the process variables" (Stephen Baldwin, personal communication).

Challenges of Developing Effective Standards

Several challenges must be addressed in developing an effective accountability system based on standards. The reform movement, as articulated in Goals 2000 and elsewhere, rests on these basic tenets about standards and tests:

- standards should be clear but not oversimplified;
- assessments should come in multiple forms and be more closely aligned with the knowledge and skills sought than conventional tests;
- standards and assessments should be understandable, acceptable, and motivating to students, teachers, and parents; and
- the focal point should be at the state and local level, guided by voluntary models developed nationally.

What criteria should standards meet to be considered worthy of certification? One set of recommendations has been published by the Goals 3 and 4 Standards Review Technical Planning Group:[2] For national subject-specific content standards, the criterion descriptors identified by the Technical Planning Group are: world-class, important and focused, useful, reflective of broad consensus-building, balanced, accurate and sound, clear and usable, assessable, adaptable, and developmentally appropriate. For state content standards, the criterion descriptors are: as rigorous as national subject-specific standards, feasible, cumulatively adequate, encouraging of students' ability to integrate and apply knowledge and skills from various subjects, and reflective of broad state consensus-building.

Although board members generally viewed these criteria as a good starting point, several areas were felt to be in need of further refinement: What does it mean to be "world class"? What is the middle ground between "specific" and "flexible"? How finely grained are the skills and knowledge being sought? To what extent should disciplinary content standards embody the skills valued in the workplace?

o o o o

"Certainly within English studies, the standards movement is trying to think through what this discipline is all about at a particular time and place in history....I feel sometimes that the documents have suggested that this job is a much simpler one than it actually is."

Miles Myers

o o o o

2 See "Promises to Keep: Creating High Standards for American Students," report to the National Education Goals Panel, November 1993, pp. iii-iv.

One challenge receiving scant attention in the popular discussion is the need to integrate standards and assessments across the various disciplines into a feasible whole and to control the proliferation of standards. There may be a temptation for professional associations to produce standards to gain visibility, resulting in multiple standards that do not mesh and, once established, are difficult to revise. It was suggested that political mechanisms be designed to address this potential problem.

° ° ° °

"Standards sound very progressive. But in their fully formed state, standards are an incredibly conservative policy instrument. What we'll be facing in a decade or so are standards that are really the congealed residue of interest-based politics around disciplines—which are going to be incredibly hard to change and incredibly difficult to ration, unless we have some mechanisms in place for questioning."

Richard Elmore

° ° ° °

Another perplexing question is how to ensure that standards are genuine motivators for improved teaching and learning. The prevailing wisdom is that content and performance standards will motivate higher performance by providing a clearer direction to schools about instructional changes needed, a clearer message to students, teachers, and parents about the performance expected, and a clearer yardstick for the public and policy makers about the progress made. If still greater motivation is desired, then higher stakes can be attached to performance in the form of sanctions and rewards.[3]

This dynamic may be more complex than prevailing wisdom assumes, however, as explained below in the section on developing aligned assessments. Some board members submitted that genuine motivation occurs only when standards are "hard currency": reflective of something meaningful in the real world, such as skills and performances valued in the international marketplace. Others felt that the evidence was fuzzy about what really motivates students.

Ensuring equity for special groups of students within a standards-based framework is another major challenge. Many feel that applying performance standards to the current system could make fiscal and other inequities more glaring; when sanctions

3 *High stakes* has become a general way of describing the use of test results to make decisions or allocate resources in ways that can have significant consequences. But the question is often "High stakes for whom?" Depending on the test and its uses, the answer can be (a) the student or test-taker, as in the case of grade retention decisions or college admissions; (b) the teacher, as in the case of using student test results as a basis for teachers' promotions or salary determinations; (c) schools or districts, as in the case of test results being reported in the newspaper or publicized in real estate advertisements; (d) states, as in the case of test results being used to rank state educational performance; (e) the nation, as in the case of national educational progress being ranked alongside performance in other countries; or (f) all (or some combination) of the above.

and rewards are attached, existing inequities could be exacerbated.

Concern was voiced that many states are making only token attempts to address key equity questions, especially in terms of fiscal equalization on the input side. A counterargument was that resources (virtually of any amount) can be used in widely different ways and that there is no assurance that new input requirements will promote greater equity or more effective use of resources.

Legal Ramifications of Standards

Will opportunity-to-learn standards generate a spate of lawsuits by parents and others dissatisfied with schools, as some have suggested? David Tatel's presentation, and the discussion that ensued, shed light on a legal aspect of reform that is often overlooked: opportunity-to-learn standards may be a less effective tool for courts to order change than content and performance standards. Courts have focused for decades on whether schools are providing inputs, Tatel explained, particularly in school finance and school desegregation cases. Opportunity-to-learn standards do not represent a departure from this approach and therefore may not significantly increase the amount of school litigation.

The adoption of state content and performance standards, however, may hasten a new trend among courts to examine student outcomes and order outcome-based remedies. Tatel noted that content and performance standards present courts with refined, ready-made tools for assessing the quality of school systems by the state's own definitions. This does not mean that courts will abandon interest in inputs and resources altogether. Rather, the typical court order in a school finance case may include outcome and input factors.

Court challenges are also like to arise from the application of new standards-based assessments, especially if the assess-

° ° ° °

"Standards are being seen as the rabbit at a greyhound track—if only we put standards out there people will chase after them, and if we make the stakes really high, people will chase even faster. What we're really trying to build is the fattest, lushest looking rabbit that can be zipped down the track to get students and teachers chasing after it."

Alan Lesgold

° ° ° °

"What will be of [particular] interest to the courts in Goals 2000 and in Chapter 1 is not so much the opportunity-to-learn standards—[although] they will be of [some] interest—but rather the content standards, the performance standards, and the assessment system designed to measure them."

David Tatel

° ° ° °

ments have high stakes or produce adverse effects for particular racial or ethnic groups. Whether courts will have confidence in the assessments—or for that matter in the standards—may depend on whether educators have confidence in them, since judges often rely on expert witnesses to illuminate complex technical issues. If the experts disagree deeply, then courts will be less likely to embrace standards and assessments.

o o o o

"The meaningfulness of content and performance standards is questionable if improved learning does not occur among the traditionally underserved."

Sylvia Johnson

o o o o

FOR FURTHER ANALYSIS

- Which applications of content and performance standards, opportunity-to-learn standards, and other governance strategies or requirements can ensure both high performance and equitable resources for learning?
- Which kinds of classroom inputs translate into desirable student outcomes?
- How can input measures be employed as part of opportunity-to-learn standards?
- Under what conditions can standards become effective motivators for students, teachers, and others?
- What should be done to ensure fair and accurate portrayals of districts, schools, and students?
- What should be done about districts, schools, and students who do not meet expected levels of progress or performance?

DEVELOPING ASSESSMENTS ALIGNED WITH STANDARDS

THE ISSUE IN BRIEF

Assessments aligned with standards are a keystone of the new reform agenda. It might be said that much of the success of standards-based reform hinges on assessments that are not yet perfected or, in some cases, even invented.

There is widespread belief that these assessments should include some type of performance measurement, given the knowledge and skills being addressed in content standards. (For example, it is difficult to test a student's knowledge of—and ability to conduct or participate in—scientific inquiry solely on the basis of multiple-choice items.) Test developers, researchers, and practitioners are already piloting various performance-based formats—portfolios of student work, written essays, observations of student performance, for instance—but many of these assessments are still in the early stages, and their effects, good or bad, are not fully known.

o o o o

"A lot of the trouble that we've gotten into on assessments is that they've been used for purposes for which they weren't designed."

Gordon Ambach

o o o o

The tendency in American education has been to apply relatively sophisticated tests to a variety of functions, including some for which they were never designed, then worry later about whether the uses were appropriate and how they affected instruction and students.

The current situation presents an opportunity for the nation to do things differently this time, by analyzing important reliability and validity questions up front, by designing standards and assessments with specific uses in mind, and by applying them cautiously to high stakes decisions. Although some reform advocates warn that an overcautious requirement of scientific rigor will delay implementation and progress, workshop participants generally agreed that a consensus is growing for careful attention to the scientific and technological bases for assessments in their various applications.

How should states approach the task of developing new assessments? What lessons can be learned from current state programs of performance-based assessment? What are the major technical considerations? How can states ensure that the new assessments are used appropriately and have a positive impact on instruction? Workshop participants weighed these and related questions.

VIEWS FROM THE WORKSHOP

Approaching Assessment Development

The enactment of Goals 2000 and the near-completion of legislation to reauthorize the Elementary and Secondary Education Act (ESEA)[4] speak to the need for an immediate and extensive research and development effort. The workshop yielded several suggestions for how a development effort could be approached.

Discussants noted that some potential pitfalls could be avoided if standard-setting groups considered assessment issues at the same time they developed content and performance standards: standards would be less likely to be built around unrealistic assumptions about what assessment technology can deliver, and federal and state governments would be less likely to attach high stakes to assessments before they were technically ready—or at least would be more aware of the consequences if they did.

In developing assessments, states would be well advised to initiate an open dialogue about the broader social and policy implications of assessment, including appropriate test use, appropriate reporting and interpretation of results, impacts on various groups of students by race, ethnicity, gender, and socioeconomic background, effects on instruction, costs and benefits of new assessments, and teacher professional development needs emanating from new standards and related assessment formats. These questions are too important to be decided by default, it was argued, and should not be dropped into the laps of test designers and measurement specialists without a public airing.

Participants strongly urged that research on assessment be a continuous process that does not end when new assessments are implemented. The process should include initial empirical research during the standards and assessment development phase, pilots and demonstrations during the preimplementation phase, and ongoing studies to monitor the implementation of the standards and assessments themselves and provide feedback for continuous revision. These studies—which might be in the form of an annual state report card on standards and assessments—could also identify areas in which additional research is needed.

o o o o

"The evidence is building that innovative assessments can be a powerful tool for reform, but it is unambiguously the case that many of the proponents have egregiously overpromised."

Daniel Koretz

o o o o

4 ESEA passed in October 1994, as the "Improving America's Schools Act of 1994." Workshop participants discussed versions of the bill as they existed in March.

Lessons from Vermont

Research should begin by studying the lessons emerging from existing innovative assessment programs. One such program is Vermont's new assessment system, which emphasizes student portfolios. The Vermont portfolio program appears to be having powerful and positive effects on instruction, according to Daniel Koretz, such as encouraging mathematics teachers to devote more time to problem solving and motivating teachers who had seemed impervious to change. But these positive effects have come with a steep price of time, stress, and money: teachers reported spending an average of 30 hours per month on portfolios, excluding training (although most say they consider the time a worthwhile burden). And from early accounts, the costs of scoring, training, and other administrative functions are likely to be much higher than the $33 per student estimated by the U.S. General Accounting Office.[5]

Preliminary evidence from Vermont raises serious questions of reliability, validity, feasibility, and bias that need more attention before portfolio data are applied on a larger scale or for high-stakes decisions, Koretz said. Scores to date have been too unreliable to be used for making comparisons across schools, for example. Efforts to appraise validity have been hindered by a lack of comparable achievement data, and the comparisons made thus far raise doubts about whether validity problems can be overcome. Teachers vary widely in their implementation of the portfolio program, which could threaten the validity of any comparison data.

Other problems in Vermont with national implications include difficulty in training large numbers of raters to a level of sufficient accuracy, a lack of standardization of performance tasks, and the limited ability to generalize about student knowledge from a small number of tasks.

The Vermont experience suggests that the twin goals of new assessments—to improve instruction and to yield high-quality comparative data—may not be totally reconcilable. A brief illustration: from an instructional perspective, it makes sense for teachers to vary performance tasks for students of different achievement levels so that lower-achieving students are not discouraged by constant failure; from a measurement perspective, however, it is problematic. Policy makers may have to accept lower levels of reliability as a price for using teacher-

5 *Student Testing: Current Extent and Expenditures, with Cost Estimates for a National Examination* (GAO/PEMD-93-8, January 13, 1993). Available from the U.S. Government Printing Office, Washington, D.C.

° ° ° °

"A major dilemma we face is that the technical tools at our disposal for assessment were created at a time when the field had a different sense of what constitutes knowledge and understanding. Thus, we have at our disposal a wonderful set of technical tools that deal with precisely the wrong questions. We need to develop technical tools that will help us make progress on issues related to the construction of meaningful and reliable standards."

Alan Schoenfeld

° ° ° °

developed and scored performance assessments for accountability purposes.

Expressed differently, this lesson from Vermont can be summarized in terms of the following tension that needs to be understood by policy makers: comparison across students or schools requires standardization, whereas improved learning for all students may require less standardization and the capacity to accommodate to specific learning needs that vary within and across classrooms.[6]

The Vermont experience affirms the wisdom of having modest expectations, evaluating the planned assessments, and allowing for a long experimentation period, luxuries that may not always be available.

Technical Questions

As indicated by the Vermont experience, a variety of technical issues—not the least of which are reliability and validity—should be the subject of extensive research. One issue needing further study is how to identify the tasks to be included in performance assessment. For example, although it may be easy to conceive of a real-world problem that engages thinking skills, content knowledge, and writing skills, it is more difficult to create an assessment item with these features that also meets reasonable measurement criteria: generalizability, reliability, and comparability. Limited generalizability of performance assessment tasks poses a particularly formidable barrier: Can a small number of items cover a content domain? Does successful performance on one task generalize to success on other tasks?

6 Vermont is, of course, not the only state in which tensions have mounted over the twin demands for standardized reporting of individual-level test data and instructionally valuable methods of assessment. The California Learning Assessment System (CLAS), for example, was an innovative program based on performance measures of achievement closely aligned to curriculum frameworks that had been developed over many years. CLAS ran into significant problems that were attributable, at least in part, to the conflicting demands for standardized data that provide a reliable basis for comparisons of individual achievement and assessments that are considered instructionally valuable. This tension was exacerbated by the need to hold down the costs of the performance assessment program by implementing a sampling methodology, which conflicted with demands that all children be included in what had been promoted as instructionally valuable exercises. The workshop discussion did not focus on the California experience; a board bulletin planned for the near future will address some of the salient issues in greater detail.

Still another critical issue is how to mix multiple measures into an integrated assessment system. How can information from performance assessments and more conventional tests be merged into a picture of progress at the student, school, and district levels? How can qualitative judgments be blended with quantitative data? What happens when the information is contradictory? When is matrix sampling appropriate, and when should universal testing be used?

Reporting of information raises another set of technical questions. Conventional reporting uses a "cut score" approach. Board members questioned, however, whether this approach is compatible with the intent of performance assessment. What is needed is a reporting approach that captures the richness of the performance but is also clear and understandable to students, parents, and the public. One suggestion was to use a "Consumer Reports" approach, with symbols and rankings for different skills and attributes and written comments that provide more detail on performance. Whatever the approach, it is likely to require a substantial public information effort to help parents, the media, and others understand new test scoring and reporting methods.

Other topics for additional technical research include approaches for assessing linguistic minorities; procedures for aggregating results across schools, districts, states, the nation, and even the globe; and interim policies for moving from current testing modes to new methods. The latter issue is particularly important with respect to proposed revisions to testing and evaluation requirements under Title I of the Elementary and Secondary Education Act (see also the discussion in the section on federalism).

° ° ° °

"There is a real danger of jumping to reliance on a set of measures and a technology that is not really there yet—and then we may find that it doesn't work very well, and go back to the things that had been familiar. There is this sense that the new measures are not corruptible; it was the old measures that were corruptible....We have to be careful that the extravagant promises being made around the country right now [for performance assessment] don't sow the seeds for the whole thing falling apart."

Robert Linn

° ° ° °

Appropriate Use of Assessments

An issue that merits early and full debate is the appropriate and fair use of various types of standards-based assessments. Board members recommended that new assessments be clearly differentiated, perhaps even labeled, as to whether they are

appropriate for diagnosing student progress and needs, monitoring or comparing the progress of teachers, schools, and school systems, governing the application of sanctions or rewards, or determining individual credentialing. It is also important to delineate whether tests are appropriate for individual use, aggregate use, or both. Cautions were raised about the possibility of the "corruptibility" of measures applied to high-stakes decisions.[7]

Effects on Teaching and Learning

Another critical issue is the effect of standards-based assessments on student learning. Some board members suggested that when tests have meaningful consequences, they influence student efforts to learn, teacher efforts to instruct, and parent efforts to support learning. Others contended that, although students may perform well on an assessment, it is difficult to know whether they have truly learned the underlying construct. Still others felt that when tests are aligned closely with local curriculum and classroom instructional methods and when the performance assessed involves higher-order skills, it does not matter whether one is able to disentangle the performance from the underlying construct or whether a student has been coached to higher levels of performance.

Related questions for research include whether certain types of assessments are better motivators than others and how new assessments affect learning disparities among various groups of students.

Another critical area for research is the effect of new assessments on instruction. Some board members questioned whether meaningful experiments could be designed to answer these kinds of questions when so many variables impinge on the learning environment. An alternative is an auditing or inspectorate approach that examines whether opportunities to learn

°°°°

"Instead of thinking about a single national evaluation, we would probably learn a lot more from a series of smaller research studies that would look at specific sectors of the population and try to answer the most important question: What works best for whom [and] under what circumstances?"

Luis Laosa

°°°°

7 Corruptible in this context means that the reliability or validity of the inferences drawn from an assessment are threatened by the behavior of test-takers or administerers of the tests. For example, "teaching to the test" means that teachers focus their lessons so as to raise the chances that their students will answer anticipated test items correctly, which can result in inflated test scores but not necessarily in increased learning of the underlying content or domain from which the test is meant to sample.

are actually being provided in the classroom and whether the curriculum being offered meets content standards. In addition, a series of smaller studies could address particular aspects of testing and learning.

FOR FURTHER ANALYSIS

- How can the nation ensure that assessments are used appropriately and fairly?
- Under what conditions is it appropriate to use assessments for high-stakes applications?
- How can we extract reliable and useful information from heterogeneous data elements that emerge in performance assessment?

A NEW ERA OF EDUCATION FEDERALISM

THE ISSUE IN BRIEF

Goals 2000 and the Elementary and Secondary Education Act reauthorization legislation have far-reaching implications for the federal, state, and local compact on education. Goals 2000 establishes a framework for standards-based reform, codifies eight national education goals in federal law, authorizes funding and other incentives to encourage states to adopt and implement standards, calls for participating states to develop assessments aligned with standards, and authorizes federal money to develop and evaluate new assessments. The ESEA legislation revises the testing and accountability requirements of the Chapter 1 program for disadvantaged children (renamed Title I).

Both Goals 2000 and ESEA contain reassurances about the voluntary nature of national standards, vest primary control of standards and assessments in the states, and establish a partnership between local communities and the federal government. What types of governance relationships are implied by the new legislation? What are the potential impacts of the federal government on state and local policies? The workshop spurred new thinking about these questions.

VIEWS FROM THE WORKSHOP

Title I: A Major Influence

Included in the ESEA reauthorization legislation are the outlines of a new system for testing and accountability under Title I. This system, analyzed in Phyllis McClure's presentation and ensuing discussion, would replace the current Title I testing procedures, which are based on national aggregation of norm-referenced test data and which have been criticized for promoting undesirable instructional approaches for disadvantaged children and for producing information of questionable quality and utility. National aggregation of local Title I test data would be abandoned; instead national information would be obtained from a national assessment that used a matrix sampling approach.

°°°°

"Title I is really the 800-pound gorilla that is going to drive Goals 2000. What Title I says about assessment is what school districts are going to follow."

Phyllis McClure

°°°°

In addition, the House version of the bill required states to adopt content, performance, and opportunity-to-learn standards for Title I children that are the same as those for all children and that are aligned with the Goals 2000 standards. States would also develop or adopt state assessments to measure the proficiency of Title I children in core academic subjects. These assessments would be administered at some point during grades 3-5, 6-9, and 10-12 and would provide individual student scores, as well as disaggregated results for certain subgroups. Assessment results would also be used to gauge the progress of schools and districts in helping Title I children meet performance standards. Sanctions and rewards stronger than those in current law would be tied to these evaluations.

With over $7 billion in federal dollars at stake and with three-quarters of the school districts in the nation participating, the Title I amendments may prove to be more consequential than Goals 2000 and, in effect, could set the parameters of a state's general assessment system.

Several concerns emerged from the workshop regarding the Title I amendments. One question revolved around the decision in the legislation to use the same system of standards and assessments for multiple purposes, from measuring individual student progress to enforcing institutional accountability. As an alternative, it was suggested that individual student assessments and institutional accountability were different functions requiring different measurements: for the former, schools could use multiple measures designed by teachers, and for the latter, standards-based assessments administered through matrix sampling.

The new provisions could actually increase the amount of testing attributable to Title I, it was argued, if multiple assessments are developed in all core subjects. Questions arose about how the multiple measures called for in the bill would be applied to state and local accountability decisions; whether the measures would meet reliability, validity, and other technical criteria; and whether the new assessments will be appropriate for high-stakes uses. Further questions focused on how to maintain baseline information on individual student achievement if assessments are administered only at certain grades, possibly beginning as late as grade 5. In many ways the new system could be more problematic than the one being replaced, warned presenter Michael Kean.

Other issues are whether the three-year period for developing new Title I assessments will be adequate and where the funding will come from to develop and pilot the mandated assessments and train teachers in their use.

°°°°

"In the name of reform, we are about to create a more complex, more technically problematic, more burdensome, and perhaps less useful assessment system."

Michael Kean

°°°°

Certifying Standards and Assessments

Goals 2000 establishes a new entity, the National Educational Standards and Improvement Council (NESIC), with the main responsibility for "certifying" standards and assessments—a challenging and complex task, participants said. There is no single definition or widely accepted set of criteria that make assessments certifiable or uncertifiable. Rather, experts can only analyze whether assessments meet various technical and other criteria. It was suggested (although this option is not specified in Goals 2000) that NESIC might produce a *range of judgments* about the relative strengths and weaknesses of specific assessments on different criteria.

Several governance issues are left unanswered by Goals 2000. For example, what is the standing of voluntary national standards that are authorized by federal legislation, developed by national but nonfederal panels, certified by a federally established body that includes nonfederal representatives, and offered as a model to guide state standards but not control state curricula? How will these standards affect state governance systems, especially when the states must answer to constituents and potential litigants? How will they influence local behavior from several layers removed? To what extent are state standards expected to be aligned with

° ° ° °

"All of this federal and state action is irrelevant if no one is checking what is happening in the classroom. What goes on in classrooms has been impervious to the actions of the federal and state levels a good deal of the time....I'd be worried that only those standards which are measurable will find their way into the schools and that experiences which are educational but unable to be measured get excluded—a trip to the museum gets thrown out of the curriculum because nobody knows what to expect from that....Are we going to do anything different this time to make sure that the enacted curriculum in the classroom is in fact compatible with content and performance standards?"

David Berliner

° ° ° °

voluntary national standards?

A particularly perplexing issue is how to retain sufficient flexibility for schools and teachers within a standards framework. Good teachers often make curricular and instructional decisions. But if standards are too detailed about content, effective teaching strategies that are not easily measured or do not hew closely to content standards could be squeezed out of the curriculum.

FOR FURTHER ANALYSIS

- What criteria should govern development of new state accountability systems for Title I? Should these systems be the same as those being developed under Goals 2000?
- What is the relationship between national and state standards?
- What criteria should NESIC consider in certifying standards and assessments?
- How much variation among states should be allowed in developing standards and assessments under Goals 2000?
- How can flexibility for different approaches to content and instruction be built into a standards framework at the local, state, and national levels?

STRENGTHENING STATE AND LOCAL CAPACITY

THE ISSUE IN BRIEF

Implementing standards-based reforms will require expertise at the state and local levels. Teachers will have to be prepared to teach the knowledge and skills embodied in content standards. Universities will have to be conversant with new thinking about content and performance in order to prepare teachers. State agency staff will have to be able to provide

technical assistance and monitor implementation of standards and assessments. Local school districts will have to adopt the organizational structures, curriculum and assessment support, and other conditions to enable teachers to teach to the standards and students to learn. Communities may need to conduct public awareness programs to help parents understand standards and assessments and their role in supporting their children's learning. What are the capacity implications of the new responsibilities being demanded of states and school districts? What types of governance arrangements can help states meet these responsibilities? The workshop discussion kept returning to these issues.

VIEWS FROM THE WORKSHOP

States vary widely in their capacity to carry out these ambitious reforms and their will to change. Local capacity is even more variable. Without specific attention to capacity building, states may be divided into those that are ready and able to implement standards and those that are not. The former group would probably include the states that have already embarked on ambitious standards and new assessments—ironically those least likely to need a push from the federal government.

State and federal policy makers would be well advised to consider the kinds of procedures and governance structures that will bridge the distance between standards on paper and practices in the classroom. Goals 2000 does not answer these questions. Although each state will have to construct its own capacity-building agenda, some type of national leadership or process would help nudge those states that lack the political will, funding, or expertise to begin.

New Governance Arrangements

States need to devise more creative governance models and strategies to influence local behavior but avoid the mistakes of the past, participants suggested.

° ° ° °

"The states have not waited for national standards. So we won't have one set of anything, we will have 30 sets of them. And we will be able to look at works in progress and be able to have a much richer set of experiences to draw from and lessons to be learned....States may not have the capacity to do a lot of these things, [but] they have the right to. As we are dealing with the sovereign right of states to set certain kinds of things in motion, we have to worry about capacity issues."

Shirley Malcom

° ° ° °

States might look for mechanisms that nudge policies in the direction of performance and outcome standards, while seeking more effective and less obtrusive methods of input regulation. These latter methods might include professional self-regulation, peer review, voluntary compliance with standards, and professionally organized technical assistance to low-performing schools.

o o o o

"Let me create two stereotypes of possible systems. One is the lean, mean performance machine, in which schools are straining to meet public expectations and input constraints are relaxed to free schools to find the right way to educate their kids.... The second is the Prussian model, which is captured by the phrase, "That which is not prohibited is required...." It is not clear that standards-based reform leads unerringly in one direction or the other."

Richard Elmore

o o o o

To get from here to there, it may be advisable to reduce input requirements whenever possible and bring existing regulation into conformance with standards. States would shift their focus from regulating inputs to setting performance goals for schools. State monitoring of compliance could be narrower but more intensive, limited only to those process requirements that passed strict review. Equity issues could be addressed through definitions of performance and incentives that would increase access of students to high-quality learning experiences. Schools could be evaluated according to a series of indicators and special studies, and in terms of the value added for students.

Indirect regulation might be achieved by adopting standards of good practice for instruction, assessment, and other important areas. One suggestion was to create a state board of teachers, teacher educators, and lay people to set professional practice standards and oversee teacher licensing. Other state panels might assume responsibility for developing and administering new assessments. States should make funding available for existing institutions, such as schools of education and professional organizations, to coordinate their policies around standards and implement mutually supportive changes in curriculum and practice. The idea is to change teaching by creating a climate in which good teaching thrives, rather than by controlling instruction. The best teachers could be engaged to lead a renewal effort and train others.

Under these strategies, the state would become less a regulator and more of a mobilizer, at the hub of a set of relationships with several government and quasi-governmental entities.

It was suggested that the assessment process itself can become a vehicle for professional development and capacity building. Engaging teachers in portfolio assessment, for

example, appears to be a valuable way to educate them about new instructional approaches and encourage them to integrate tasks important for students to learn.

FOR FURTHER ANALYSIS

- What types of national leadership can influence states with widely varying capacities and prevent further stratification?
- What types of supports will states need to strengthen their capacity to carry out standards-based reforms successfully?
- How can states be encouraged to implement new governance structures compatible with standards-based reforms?

° ° ° °

"No matter what anybody decides about standards and assessment procedures, the most important thing to look at is how they can be linked to the community within which they have to be used."

Brigitte Jordan

° ° ° °

NEXT STEPS

The wide range of issues covered during the workshop reflects the newness and the complexity of standards-based reform, and the discussions reflected a widespread enthusiasm for the possibilities for genuine improvement embodied in the standards-based reform movement. The possibilities for effective reform are especially exciting to many educators today in the light of new research on how children learn, what kinds of nontraditional learning environments are best suited to learners, and how teachers' understanding of the educational process can affect the development and uses of standards.[8]

Many decisions will have to be made in the near future for the vision of reform to become a reality:

- the national standards-setting committees will continue their work;
- states will continue (or begin) to implement Goals 2000;
- the U.S. Department of Education will begin to develop regulations for Title I and parameters for the National Assessment of Educational Progress; and

8 Ann Brown, personal communication, October 1994.

- the new National Skill Standards board—also established by Goals 2000—will convene and begin to evaluate and certify national standards defining knowledge and competencies required for clusters of jobs in the U.S. economy.

Throughout this process, the Board on Testing and Assessment will continue to foster dialogue and provide support and information to policy makers on standards and assessment issues. The issues and questions raised during the workshop are the beginning of a long-term systematic effort by the board to help identify and answer difficult questions. Many follow-up activities are already planned:

- The board has launched a major committee study of the effects of Goals 2000 on students with disabilities, as mandated in the act. This study, which will take two years to complete, will have important implications for the next stages of standards-based reform, especially as it affects issues of inclusion, accommodations for students with special needs, and other equity concerns.

In addition, the board is planning:

- orientation briefings and discussion meetings for federal agency;
- in-depth analysis of performance standards methods, comparison of approaches being tried in various states and/or other countries, and policy implications;
- the exploration of technical issues pertaining to implementation of Title I testing and evaluation requirements;
- the development of technical analyses and policy options regarding the status of the National Assessment of Educational Progress (NAEP) under Goals 2000;
- the development of forums for teachers to discuss their role in standards-based reform;
- the establishment of mechanisms to help the media improve the coverage of test-based information on schools and labor market performance; and
- convening of regular inter-agency discussions on links between educational and occupational skill standards issues.

° ° ° °

"It is clear that we lack the precision that a lot of people would like to have in these areas. I hope that we will recognize the lack of precision and that we are careful not to do any harm when we clearly don't understand all the problems."

Richard Atkinson

° ° ° °

APPENDIX

WORKSHOP AGENDA AND PARTICIPANTS

Standards and Assessment:
Toward an Agenda for Policy Research

A WORKSHOP OF THE BOARD ON TESTING AND ASSESSMENT

Lecture Room, National Academy of Sciences
2101 Constitution Avenue, NW
Washington, DC

March 9, 1994

Constance B. Newman, Vice-Chair, BOTA, Presiding

° ° ° °

8:00 am	Pastries and coffee
	Introduction and welcoming remarks *Suzanne Woolsey, Chief Operations Officer, NAS* *Constance Newman*
8:30	**Content and Performance: Defining Terms**
	Presentation: *Shirley Malcom* (AAAS), Chair, Goals Panel Technical Planning Group, Goals 3 and 4 Standards Review "Promises to Keep: High Standards for American Students"
	Response: *Richard Elmore* (BOTA)
	General discussion
10:00	Break
10:15	**Opportunity to Learn: Equity and Accountability**
	Perspectives: *David Tatel* (BOTA): Opportunity to Learn, Opportunities to Sue *Susan Fuhrman* (Rutgers CPRE): Lessons on the Politics of Standards
	Responses: *Sylvia Johnson* (BOTA) *Ewart Thomas* (BOTA)
	General discussion

11:45	Comments from observers and invited guests
NOON	**Lunch**
12:45 pm	**Greetings from Bruce Alberts, President, NAS**
1:00	**The New Educational Federalism: Linking Goals 2000 and ESEA** Perspectives: *Phyllis McClure* (Washington, DC): Anticipating the New Title I *Michael Kean* (CTB Macmillan/McGraw-Hill): National Norms and Local Needs Responses: *David Berliner* (BOTA) *Edmund Gordon* (BOTA) General discussion
2:30	Break
2:45	**Incentives for Individual and System Performance: The Role of Testing and Assessment** Perspectives: *Daniel Koretz* (RAND): Lessons from Vermont *Gordon Ambach* (CCSSO): The States and the Nation Responses: *Alan Schoenfeld* (BOTA) *Robert Linn* (BOTA) General discussion
4:15	Comments from observers and invited guests
4:30	**Synthesis: Outlining a Policy Research Agenda** Remarks: *Alan Lesgold* (BOTA) Closing comments: Richard Atkinson (Chair, BOTA) General discussion
5:15	**Reception**

PARTICIPANTS

GORDON AMBACH, Council of Chief State School Officers, Washington, D.C.

RICHARD C. ATKINSON, University of California, San Diego

LAURIE J. BASSI, Graduate Public Policy Program, Georgetown University, Washington, D.C.

DAVID C. BERLINER, College of Education, Arizona State University, Tempe

RICHARD F. ELMORE, Graduate School of Education, Harvard University, Cambridge

MICHAEL J. FEUER, National Research Council, Washington, D.C.

SUSAN H. FUHRMAN, Consortium for Policy Research in Education, Rutgers University

SYLVIA T. JOHNSON, School of Education, Howard University, Washington

BRIGITTE JORDAN, Xerox Palo Alto Research Center and Institute for Research on Learning, Palo Alto, Calif.

CARL F. KAESTLE, Wisconsin Center for Education Research, University of Wisconsin at Madison

MICHAEL H. KEAN, CTB McGraw-Hill, Monterey, Calif.

DANIEL KORETZ, Rand Corporation, Washington, D.C.

LUIS M. LAOSA, Educational Testing Service, Princeton, N.J.

RENEE S. LERCHE, Ford Motor Company, Dearborn, Mich.

ALAN M. LESGOLD, Learning Research and Development Center, University of Pittsburgh

ROBERT L. LINN, School of Education, University of Colorado, Boulder

SHIRLEY MALCOM, American Association for the Advancement of Science, Washington, D.C.

PHYLLIS MCCLURE, Washington, D.C.

MILES A. MYERS, National Council of Teachers of English, Urbana, Ill.

CONSTANCE B. NEWMAN, Smithsonian Institution, Washington, D.C.

JAMES L. OUTTZ, Outtz and Associates, Washington, D.C.

NEAL W. SCHMITT, Department of Psychology, Michigan State University, East Lansing

ALAN H. SCHOENFELD, School of Education, University of California, Berkeley

DAVID S. TATEL, Hogan and Hartson, Washington, D.C.

EWART A.C. THOMAS, Department of Psychology, Stanford University

ALEXANDRA K. WIGDOR, National Research Council, Washington, D.C.